

IN

# 14



PCT09

ENTERED

## RAW SEQUENCE LISTING

DATE: 10/24/2002

PATENT APPLICATION: US/09/868,328A

TIME: 16:16:10

Input Set : A:\EP.txt

Output Set: N:\CRF4\10242002\I868328A.raw

3 <110> APPLICANT: RHEE, Sangki  
4 SONG, Kibang  
5 KIM, Chulho  
6 RYU, Eunja  
7 LEE, Yongbok  
9 <120> TITLE OF INVENTION: Enzymatic Production of Difructose Dianhydride IV from  
sucrose and  
10 relevant enzymes and genes coding for them  
12 <130> FILE REFERENCE: 24679  
14 <140> CURRENT APPLICATION NUMBER: US 09/868,328A  
15 <141> CURRENT FILING DATE: 2001-06-18  
17 <150> PRIOR APPLICATION NUMBER: PCT/KR00/01183  
18 <151> PRIOR FILING DATE: 2000-10-19  
20 <160> NUMBER OF SEQ ID NOS: (4) ✓  
22 <170> SOFTWARE: PatentIn version 3.1  
24 <210> SEQ ID NO: 1  
25 <211> LENGTH: 2001  
26 <212> TYPE: DNA  
27 <213> ORGANISM: Arthrobacter ureafaciens K2032  
29 <400> SEQUENCE: 1  
30 gcggtgcacc ccgacttccc ctgcagcacc accgtccccc taccggccga ccgcccggcc 60  
32 cgactgctcc tcagcctaga cgggcccctc ctcgaggtct tcgtcgggga cggtgaggcg 120  
34 actgcgtoga acctggtcct cctggggggc ggcggtgtga ccgcgagcct cgagacggca 180  
36 cggccaggaa ccgtgcacgt gaccgcgacg gacgtcgagg cgcgccagca tgctgacgcc 240  
38 cctgaacctg ccgcggttct gggtgacga gcgtccccc cccgacagct ctctttctac 300  
40 cgctgcccga accaggggtg acgcttcgtc gcgcccaccc gtccacgaga ggaaccagca 360  
42 atgacgcggg ccattctcag ccgcgcgctg ctccagggag ccggcgccgg agcaactgcc 420  
44 ctgatcttcg gcggtgctgt gccgcctgca gcccgggcat ccgctccggg ctgctccgct 480  
46 gccgtctacc acatgacgcc cccagcggc tggtcttgcg acccccaacg cccggtcacc 540  
48 acccacggcg cctaccagct gtactacctg cactccgacc agaacaacgg ccccgcgggc 600  
50 tgggaccacg cgagcagcag cgacggcgct gccttcacgc accacggcac cgtgatgccg 660  
52 ctgcggcccg acttcccctg gtggtccggg tcggcggtcg tcggcaccgc gaacacggca 720  
54 ggggttcggcg ccggcgcggt cgtcgcgctc gcgaccagc cgaccgacgg cgtccgcaag 780  
56 taccaggagc agtacctcta ctggtcgacc gacggcggtt tcacgttcac cgccctgcc 840  
58 gaccccgta tcgtcaaac cgacggctgc gccgccacca cgcgcgcga gatcgagaac 900  
60 gccgagtggg tcgcgacccc caagatccac tgggacaccg cccgcggaga atgggtctgc 960  
62 gtcacgggac gactgcggta ccgcgcgttc tacacctgc cgaacctgcg cgactggaca 1020  
64 ctgcgcgca acttcgacta cccgaaccac gccctcggcg gcacgcagtg ccccgacctg 1080  
66 ttgcagatca ccgcagacga cgggacacgc cactgggtgc tcgcgcgag catggacgcc 1140  
68 tacggcatcg gcttccccat gacgtacgcc tactggacag gcacctggga cggcgagcag 1200  
70 ttccacgcgg acgacctcac cccgcaatgg ctgcactggg gctgggactg gtacgcggcc 1260  
72 gtcacctggc catcgatcga cgcgcccag accaagcgcc tcgccatcgc gtggatgaac 1320  
74 aactggaagt acgcgcgacg cgacgtcccc accgacgcat ccgacggcta caacgggcag 1380  
76 aactcgatcg tccgcgagct gcggctcgcc cgacagcctg gcggctggta caccctcctg 1440

## RAW SEQUENCE LISTING

DATE: 10/24/2002

PATENT APPLICATION: US/09/868,328A

TIME: 16:16:10

Input Set : A:\EP.txt

Output Set: N:\CRF4\10242002\I868328A.raw

```

78 agcaccctcg tggcagcgct gacgaactac gtcaccgcca ccaccacact ccccgaccgg 1500
80 accgtcgacg gcagcgccgt cctgccatgg aacggacgcg catacgagat cgagctcgac 1560
82 atcgctggg acaccgacgac gaacgtcggc atctcggtgg gccgctcccc cgacggaacc 1620
84 cggcacacga acatcgga gtagggagca gacctgtacg tcgaccgagg accctccgac 1680
86 ctgcgcgggt actcgctcgc cccctactcg cgagccgccc ccccatcga cccggcgcc 1740
88 cgatccgtgc acctgcgat cctcgtcgac acccagagcg tcgaggctct cgtcaacgcc 1800
90 ggccacacgg tgcctctcca gcaggtccac ttgcgcgagg gcgacacggg aatctcgctc 1860
92 tacaccgacg ggggccccgc acatttcacc ggcctcgtcg tccgcgagat tggccaggcg 1920
94 atctaggcga tgcacaccac accgctcacc gaagccgcgc cccgggagac gacggccgac 1980
96 aatcgacacg tcctcgtcgt t 2001
99 <210> SEQ ID NO: 2
100 <211> LENGTH: 521
101 <212> TYPE: PRT
102 <213> ORGANISM: Arthrobacter ureafaciens
104 <400> SEQUENCE: 2
106 Met Thr Pro Ala Ile Ser Arg Arg Ala Val Leu Gln Gly Ala Gly Ala
107 1 5 10 15
110 Gly Ala Leu Ala Leu Ile Phe Gly Gly Ala Val Pro Pro Ala Ala Arg
111 20 25 30
114 Ala Ser Ala Pro Gly Ser Leu Arg Ala Val Tyr His Met Thr Pro Pro
115 35 40 45
118 Ser Gly Trp Leu Cys Asp Pro Gln Arg Pro Val Thr Thr His Gly Ala
119 50 55 60
122 Tyr Gln Leu Tyr Tyr Leu His Ser Asp Gln Asn Asn Gly Pro Gly Gly
123 65 70 75 80
126 Trp Asp His Ala Ser Thr Thr Asp Gly Val Ala Phe Thr His His Gly
127 85 90 95
130 Thr Val Met Pro Leu Arg Pro Asp Phe Pro Val Trp Ser Gly Ser Ala
131 100 105 110
134 Val Val Gly Thr Ala Asn Thr Ala Gly Phe Gly Ala Gly Ala Val Val
135 115 120 125
138 Ala Leu Ala Thr Gln Pro Thr Asp Gly Val Arg Lys Tyr Gln Glu Gln
139 130 135 140
142 Tyr Leu Tyr Trp Ser Thr Asp Gly Gly Phe Thr Phe Thr Ala Leu Pro
143 145 150 155 160
146 Asp Pro Val Ile Val Asn Thr Asp Gly Arg Ala Ala Thr Thr Pro Ala
147 165 170 175
150 Glu Ile Glu Asn Ala Glu Trp Phe Arg Asp Pro Lys Ile His Trp Asp
151 180 185 190
154 Thr Ala Arg Gly Glu Trp Val Cys Val Ile Gly Arg Leu Arg Tyr Ala
155 195 200 205
158 Ala Phe Tyr Thr Ser Pro Asn Leu Arg Asp Trp Thr Leu Arg Arg Asn
159 210 215 220
162 Phe Asp Tyr Pro Asn His Ala Leu Gly Gly Ile Glu Cys Pro Asp Leu
163 225 230 235 240
166 Phe Glu Ile Thr Ala Asp Asp Gly Thr Arg His Trp Val Leu Ala Ala
167 245 250 255
170 Ser Met Asp Ala Tyr Gly Ile Gly Leu Pro Met Thr Tyr Ala Tyr Trp
171 260 265 270

```

## RAW SEQUENCE LISTING

DATE: 10/24/2002

PATENT APPLICATION: US/09/868,328A

TIME: 16:16:11

Input Set : A:\EP.txt

Output Set: N:\CRF4\10242002\I868328A.raw

```

174 Thr Gly Thr Trp Asp Gly Glu Gln Phe His Ala Asp Asp Leu Thr Pro
175           275           280           285
178 Gln Trp Leu Asp Trp Gly Trp Asp Trp Tyr Ala Ala Val Thr Trp Pro
179           290           295           300
182 Ser Ile Asp Ala Pro Glu Thr Lys Arg Leu Ala Ile Ala Trp Met Asn
183 305           310           315           320
186 Asn Trp Lys Tyr Ala Ala Arg Asp Val Pro Thr Asp Ala Ser Asp Gly
187           325           330           335
190 Tyr Asn Gly Gln Asn Ser Ile Val Arg Glu Leu Arg Leu Ala Arg Gln
191           340           345           350
194 Pro Gly Gly Trp Tyr Thr Leu Leu Ser Thr Pro Val Ala Ala Leu Thr
195           355           360           365
198 Asn Tyr Val Thr Ala Thr Thr Thr Leu Pro Asp Arg Thr Val Asp Gly
199           370           375           380
202 Ser Ala Val Leu Pro Trp Asn Gly Arg Ala Tyr Glu Ile Glu Leu Asp
203 385           390           395           400
206 Ile Ala Trp Asp Thr Ala Thr Asn Val Gly Ile Ser Val Gly Arg Ser
207           405           410           415
210 Pro Asp Gly Thr Arg His Thr Asn Ile Gly Lys Tyr Gly Ala Asp Leu
211           420           425           430
214 Tyr Val Asp Arg Gly Pro Ser Asp Leu Ala Gly Tyr Ser Leu Ala Pro
215           435           440           445
218 Tyr Ser Arg Ala Ala Ala Pro Ile Asp Pro Gly Ala Arg Ser Val His
219           450           455           460
222 Leu Arg Ile Leu Val Asp Thr Gln Ser Val Glu Val Phe Val Asn Ala
223 465           470           475           480
226 Gly His Thr Val Leu Ser Gln Gln Val His Phe Ala Glu Gly Asp Thr
227           485           490           495
230 Gly Ile Ser Leu Tyr Thr Asp Gly Gly Pro Ala His Phe Thr Gly Ile
231           500           505           510
234 Val Val Arg Glu Ile Gly Gln Ala Ile
235           515           520

```

238 &lt;210&gt; SEQ ID NO: 3

239 &lt;211&gt; LENGTH: 1800

240 &lt;212&gt; TYPE: DNA

241 &lt;213&gt; ORGANISM: Zymomonas mobilis

243 &lt;400&gt; SEQUENCE: 3

```

244 aaatataaat ataaatttag ttttaagtaa aaacttccga aaaaaatgtg acttcgcttg      60
246 aaatgtcgaa gcgtgtcttt cgaaaaggca gcctatttca atattttctgc ctttttgaaa      120
248 gtcgagaaca agaattaatc ttttttcggc tttatttcat cattttattag gatagttott      180
250 atgttgaata aagcaggcat tgcagagccg agcttggtga ctcgtgcgga tgctatgaaa      240
252 gtgcataacc atgatccac ggcaaccatg cctaccattg attatgactt tcctgtcatg      300
254 actgataaat attgggtttg ggacacttgg cccttacgcg atattaacgg tcaggttgct      360
256 agcttccaag gttggtcggt gatctttgct ttggtcgctg atgcaccaa atatggttgg      420
258 cataatcgca atgatggcg cagaattggt tatttctatt cacgtggtgg aagcaactgg      480
260 atttttggtg gtcattctt gaaagatggt gccaatccgc gttcttggga atggtctggt      540
262 tgcacgatta tggcaccggg tacggccaat tctgtogaag tattctttac gtctgtcaat      600
264 gatacgccgt cagaatccgt tcctgcccag tgcaagggtc acatctatgc cgatgataaa      660
266 tcggtatggt ttgacggttt tgataaagtg accgatctgt ttcaggcaga tggcctttat      720

```

## RAW SEQUENCE LISTING

DATE: 10/24/2002

PATENT APPLICATION: US/09/868,328A

TIME: 16:16:11

Input Set : A:\EP.txt

Output Set: N:\CRF4\10242002\I868328A.raw

```

268 tatgctgatt atgcagaaaa taatttctgg gatttccgcg atccgcatgt cttcattacc 780
270 ccgaagatag gcaaaacata tgccttggtt gaaggtaatg ttgccatgga gcgcggtacg 840
272 gtcgctgttg gcgaagagga aattggccct gttccaccaa aaaccgaaac gcctgatggc 900
274 gctcgctatt gtgctgctgc cattggtatt gcacaggccc ttaatgaagc ccgcaccgaa 960
276 tggaaattgt taccgccttt ggtaaccgcc tttggtgtca atgaccagac ggagcggcct 1020
278 catgtcgttt tccagaatgg cttgacctat ctctttacga tcagtcatca ttcgacttat 1080
280 gccgatgggt tgcggggtcc tgatgggggt tatggctttg tttctgaaaa cggcattttt 1140
282 gggccttatg aaccgctgaa tggttccggt ttggttctcg gtaacccttc ttcacagcct 1200
284 tatcaggctt attccatta tgtgatgaca aatgggctgg tgacctcctt cattgatacc 1260
286 attccgagtt ctgaccgaa tgtctatcgt tatggtggca ccttggcacc gaccatcaaa 1320
288 ttggaattgg ttggccatcg cagcttcggt accgaagtga agggttatgg ctatatccg 1380
290 ccacagatcg agtgggtggc agaagatgaa tcttctaatt ctgcggcagc cctgtcttta 1440
292 ttgaataaat aagatttatt cattgtaaat gacgtttctg attatgcttg aataaagcat 1500
294 tatcatcttt tgagtttatt tcaacttaag ggtcatcagg gcattttgcc ctgatgacac 1560
296 ccttctctatt cctaataata atttcaagaa agtgtatatt actttaaatg ttttaattta 1620
298 atgccagtcg ctggacgcga gcccaagcga tgaaagtgaa taaatttgat ttgacgacct 1680
300 ctatgccgga aatcggcact gattttccca ttatgcgtga tgacttgtgg ctgtgggata 1740
302 cttggccatt acgggatatc aatggcaatc ctgtcagctt taaaggctgg aatgttattt 1800
305 <210> SEQ ID NO: 4
306 <211> LENGTH: 423
307 <212> TYPE: PRT
308 <213> ORGANISM: Zymomonas mobilis
310 <400> SEQUENCE: 4
312 Met Leu Asn Lys Ala Gly Ile Ala Glu Pro Ser Leu Trp Thr Arg Ala
313 1 5 10 15
316 Asp Ala Met Lys Val His Thr Asp Asp Pro Thr Ala Thr Met Pro Thr
317 20 25 30
320 Ile Asp Tyr Asp Phe Pro Val Met Thr Asp Lys Tyr Trp Val Trp Asp
321 35 40 45
324 Thr Trp Pro Leu Arg Asp Ile Asn Gly Gln Val Val Ser Phe Gln Gly
325 50 55 60
328 Trp Ser Val Ile Phe Ala Leu Val Ala Asp Arg Thr Lys Tyr Gly Trp
329 65 70 75 80
332 His Asn Arg Asn Asp Gly Ala Arg Ile Gly Tyr Phe Tyr Ser Arg Gly
333 85 90 95
336 Gly Ser Asn Trp Ile Phe Gly Gly His Leu Leu Lys Asp Gly Ala Asn
337 100 105 110
340 Pro Arg Ser Trp Glu Trp Ser Gly Cys Thr Ile Met Ala Pro Gly Thr
341 115 120 125
344 Ala Asn Ser Val Glu Val Phe Phe Thr Ser Val Asn Asp Thr Pro Ser
345 130 135 140
348 Glu Ser Val Pro Ala Gln Cys Lys Gly Tyr Ile Tyr Ala Asp Asp Lys
349 145 150 155 160
352 Ser Val Trp Phe Asp Gly Phe Asp Lys Val Thr Asp Leu Phe Gln Ala
353 165 170 175
356 Asp Gly Leu Tyr Tyr Ala Asp Tyr Ala Glu Asn Asn Phe Trp Asp Phe
357 180 185 190
360 Arg Asp Pro His Val Phe Ile Thr Pro Lys Ile Gly Lys Thr Tyr Ala
361 195 200 205

```

## RAW SEQUENCE LISTING

DATE: 10/24/2002

PATENT APPLICATION: US/09/868,328A

TIME: 16:16:11

Input Set : A:\EP.txt

Output Set: N:\CRF4\10242002\I868328A.raw

```

364 Leu Phe Glu Gly Asn Val Ala Met Glu Arg Gly Thr Val Ala Val Gly
365      210      215      220
368 Glu Glu Glu Ile Gly Pro Val Pro Pro Lys Thr Glu Thr Pro Asp Gly
369 225      230      235      240
372 Ala Arg Tyr Cys Ala Ala Ala Ile Gly Ile Ala Gln Ala Leu Asn Glu
373      245      250      255
376 Ala Arg Thr Glu Trp Lys Leu Leu Pro Pro Leu Val Thr Ala Phe Gly
377      260      265      270
380 Val Asn Asp Gln Thr Glu Arg Pro His Val Val Phe Gln Asn Gly Leu
381      275      280      285
384 Thr Tyr Leu Phe Thr Ile Ser His His Ser Thr Tyr Ala Asp Gly Leu
385      290      295      300
388 Ser Gly Pro Asp Gly Val Tyr Gly Phe Val Ser Glu Asn Gly Ile Phe
389 305      310      315      320
392 Gly Pro Tyr Glu Pro Leu Asn Gly Ser Gly Leu Val Leu Gly Asn Pro
393      325      330      335
396 Ser Ser Gln Pro Tyr Gln Ala Tyr Ser His Tyr Val Met Thr Asn Gly
397      340      345      350
400 Leu Val Thr Ser Phe Ile Asp Thr Ile Pro Ser Ser Asp Pro Asn Val
401      355      360      365
404 Tyr Arg Tyr Gly Gly Thr Leu Ala Pro Thr Ile Lys Leu Glu Leu Val
405      370      375      380
408 Gly His Arg Ser Phe Val Thr Glu Val Lys Gly Tyr Gly Tyr Ile Pro
409 385      390      395      400
412 Pro Gln Ile Glu Trp Leu Ala Glu Asp Glu Ser Ser Asn Ser Ala Ala
413      405      410      415
416 Ala Leu Ser Leu Leu Asn Lys
417      420

```

VERIFICATION SUMMARY

DATE: 10/24/2002

PATENT APPLICATION: US/09/868,328A

TIME: 16:16:12

Input Set : A:\EP.txt

Output Set: N:\CRF4\10242002\I868328A.raw